AD_____

Award Number: DAMD17-02-1-0411

TITLE: Novel Membrane-Associated Targets for Diagnosis and Treatment of Breast Cancer

PRINCIPAL INVESTIGATOR: Brenton G. Mar

CONTRACTING ORGANIZATION: University of Illinois
Champaign, IL 61820-6242

REPORT DATE: May 2005

TYPE OF REPORT: Annual Summary

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 074-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	May 2005	Annual Summary (1 May 2002 - 30 Apr 2005)

4. TITLE AND SUBTITLE
Novel Membrane-Associated Targets for Diagnosis and Treatment of Breast Cancer

5. FUNDING NUMBERS
DAMD17-02-1-0411

6. AUTHOR(S)
Brenton G. Mar

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Illinois
Champaign, IL 61820-6242

E-Mail: Bmar1@uic.edu

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSORING / MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 Words)

Differentially expressed proteins localized to the cell membrane or secreted show great promise as therapeutic targets and diagnostic markers because of their easy accessibility. However, determining protein localization by traditional methods is a difficult process. We predict protein localization in the MCF7 breast cancer cell line by analysis of differential hybridization levels of RNAs that have been physically separated with a sucrose density gradient by virtue of their association with polysomes on the endoplasmic reticulum. Assignment to membrane-associated and secreted class membership is based on both the differential hybridization levels and an expression threshold, which are calculated empirically from data collected on a reference set of known cytoplasmic and membrane proteins. Applied to the unknown set, these criteria identified 755 probe sets as potentially MS for which this annotation did not previously exist. The data were used to filter a previously reported expression dataset to identify MS genes which are associated with poor prognosis in breast cancer and represent potential targets for diagnosis and treatment. This approach provides a useful tool for the analysis of gene expression patterns, to identify genes encoding membrane-associated or secreted proteins with biological relevance that have the potential for clinical applications in diagnosis or treatment.

14. SUBJECT TERMS
Affymetrix GeneChip, microarray, membrane-bound polysomes, amplicon, biomarkers, gene discovery

15. NUMBER OF PAGES
7

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

Table of Contents

Task 2 Determine the predictive ability of this data set against both known membrane-bound and cytoplasmic proteins, and generate an annotated database of genes encoding proteins likely to be membrane-bound or secreted in MCF7 cells (Months 13-24):

a. Completed during last reporting period.
b. Completed during last reporting period.
c. Additional data such as cytogenetic position, UniGene cluster number, and protein homology will be collected on each transcript. At this stage, we will generate an annotated database of genes encoding proteins likely to be membrane-bound or secreted in MCF7 cells. An annotated database of genes encoding cytosolic proteins will be generated as well (Month 20-24).

A database of genes encoding proteins known or predicted to be membrane bound or secreted (MS genes) in MCF7 cells (MCF-7 MS gene dataset) was generated which included 531 known and 810 predicted MS genes. Predicted MS genes in MCF7 cells met two criteria: 1) a minimal total expression level of 738 which corresponds to 24.6% of the most highly expressed Affymetrix probesets in MCF7 cells and 2) a MS/CYT ratio of 1.08 or above which indicates an enrichment in the membrane bound polysome fraction. These two criteria were selected empirically to achieve a reasonable sensitivity (80.7%) and excellent specificity (96.9%).

The MS/CYT ratio threshold of 1.08 was set to almost maximize specificity with a reasonable sensitivity of identifying MS genes. Because a significant number of MS genes in the training set had low MS/CYT ratios that overlapped with genes encoding cytoplasmic proteins, it was not possible to generate a database of predicted cytoplasmic proteins with high specificity or generate a database of very high sensitivity while maintaining a high specificity. MS genes may have low MS/CYT ratios for several reasons, including alternate mechanisms of membrane targeting, cytoplasmic export, and dissociation from the rough endoplasmic reticulum during processing.

Annotation MS genes was done in an automated fashion with information from the Unigene and the Gene Ontology database, including information on gene location, cytoband location, Unigene cluster number, protein homology, and cellular function, if available.

Task 3. Identify genes encoding membrane-bound and secreted proteins that are known to be amplified, overexpressed, or differentially expressed in breast cancer. (Months 25-36):

a. Use data from Task *2* to predict genes encoding membrane-bound and secreted proteins from amplicon data being generated in the mentor's lab from "genomic microarrays". Collect data from the literature supporting these candidates as potential drug targets and markers (Months 25-28).

In order to identify novel regions of genomic amplification in breast cancer, the lab obtained novel breast cancer cell lines established from patient biopsies. As part of another project, genomic microarrays from Vysis corp and Spectral Genomics were hybridized against these novel breast cancer lines. Unfortunately, due to technical limitations, this project did not yield any novel amplicons to analyze. Had a novel amplicon been identified, the MS database would have been used to identify genes encoding MS proteins to focus future studies on.

b. Use data from Task 2 to predict genes encoding membrane-bound and secreted proteins from candidates identified in the literature. Collect data from the literature supporting these candidates as potential drug targets and markers (Months 29-32).

The MCF-7 MS gene dataset was used to identify potential MS genes in a differential gene expression study in breast cancer which compared tumors with good vs. poor 5-year outcome [1]. Identifying MS genes may facilitate the selection of target genes for further evaluation.

In the van't Veer study, RNA from 98 primary breast tumors was hybridized to cDNA microarrays, and the resultant analysis led to a 231-gene expression profile associated with poor prognosis. The original study was preformed on cDNA glass slide microarrays; we therefore needed to find which elements of the Affymetrix U133A microarray corresponded to the 231 genes from the original study. It was possible to map 166 of these 231 genes to 269 probe sets on the Affymetrix microarray. Of these 269 probe sets, 20 were found in our predicted MS database representing 15 unique genes (see Table 1); an additional 52 were found in our training set of previously known MS genes. Of the genes not in the training set, almost half (7 out of 15) had no subcellular location annotation in GO or SwissProt, although one had a published characterization. Out of the 9 genes with functional annotation, five are involved in metabolism, along with one each involved in signal transduction, cell-cycle regulation, proteolysis, and calcium binding. It is interesting to note that of the genes without functional annotation, HCCR1 is a putative proto-oncogene, fucosyltransferase 8 is thought to contribute to malignancy, "G protein-coupled receptor 126" contains a "protein tyrosine phosphatase-like protein" domain, and "hypothetical protein FLJ22341" contains a rhomboid domain, thought to regulate epidermal growth factor receptor expression. Any of these proteins, whose upregulation is associated with poor prognosis in breast cancer, merit further investigation as potential treatment targets.

Table 1

Affymetrix ID	Original Accession #	Gene Name	Description	Localization annotation (GO and SwissProt)	MS/CYT Ratio
212640_at	AF052159		Homo sapiens clone 24416 mRNA sequence	None	1.294
212248_at	AK000745		Homo sapiens cDNA FLJ20738 fis, clone HEP08257	None	1.261
212250_at	AK000745		Homo sapiens cDNA FLJ20738 fis, clone HEP08257	None	1.232
212251_at	AK000745		Homo sapiens cDNA FLJ20738 fis, clone HEP08257	None	1.217
201818_at	AF052162	FLJ12443	hypothetical protein FLJ12443	None	1.205
218686_s_at	Contig55188_RC	FLJ22341	hypothetical protein FLJ22341	None	1.116
219202_at	Contig55188_RC	FLJ22341	hypothetical protein FLJ22341	None	1.133
207170_s_at	NM_015416	HCCR1	cervical cancer 1 protooncogene	None	1.080
201037_at	D25328	PFKP	phosphofructokinase, platelet	None	1.115
219197_s_at	NM_020974	CEGP1	CEGP1 protein	Not annotated, but literature suggests secreted protein	1.327
208658_at	NM_004911	ERP70	protein disulfide isomerase related protein (calcium-binding protein, intestinal-related)	Endoplasmic reticulum	1.221
211048_s_at	NM_004911	ERP70	protein disulfide isomerase related protein (calcium-binding protein, intestinal-related)	Endoplasmic reticulum	1.263
210074_at	NM_001333	CTSL2	cathepsin L2	Lysosome	1.310
212290_at	AL050021		Homo sapiens mRNA; cDNA DKFZp564D016 (from clone DKFZp564D016)	Membrane protein	1.212
212295_s_at	AL050021		Homo sapiens mRNA; cDNA DKFZp564D016 (from clone DKFZp564D016)	Membrane protein	1.223

213094_at	AL080079	DKFZP564D0462	hypothetical protein DKFZp564D0462	Membrane protein	1.345
219410_at	NM_018004	FLJ10134	hypothetical protein FLJ10134	Membrane protein	1.210
221675_s_at	NM_020244	LOC56994	cholinephosphotransferase 1	Membrane protein	1.356
203988_s_at	NM_004480	FUT8	fucosyltransferase 8 (alpha (1,6) fucosyltransferase)	Membrane protein (by similarity).	1.206
203362_s_at	NM_002358	MAD2L1	MAD2 (mitotic arrest deficient, yeast, homolog)-like 1	Nucleus	1.112

Table 1. **MS genes in a breast cancer expression dataset.** Genes from the 231-gene poor prognosis profile (van't Veer et al.) predicted to have MS localization are shown. Those that were found in the training set are not listed here. Accession number is shown as given in the original report; gene name and description are from GenBank.

c. Develop data into an online public resource that breast cancer researchers can use to quickly screen their candidates for membrane-bound and secreted proteins (Months 33-36).

The MCF-7 MS gene dataset is available online in excel format at the following URL:

http://www.uic.edu/~bmar1/MCF7/

Included are all probesets which meet the total expression and differential expression criteria as described above. The probesets are annotated with data from Affymetrix and other online resources and also include the total expression levels and MS/CYT ratio. Investigators can download the dataset and utilize it to identify potential MS genes within their own datasets, as demonstrated in the previous Task.

Key Accomplishments

This reporting period:

- Developed annotated dataset of genes encoding membrane bound and secreted proteins in MCF7 breast cancer cell line
- Used dataset to identify MS proteins in a published profile of genes denoting a good or poor prognosis in breast cancer
- Compiled dataset in easily accessible format and posted online for other investigators to access

Reportable Outcomes

Publication:

Stitziel NO, Mar BG, Liang J, Westbrook CA. Membrane-associated and secreted genes in breast cancer. Cancer Res. 2004 Dec 1;64(23):8682-7.

Conclusions/Summary

In summary, we have used a genome wide biological technique to identify a novel set of MS genes expressed in MCF-7 cells. MS proteins have shown great clinical utility. Membrane-bound proteins include surface antigen targets for diagnosis or treatment, such as receptors that regulate cell growth, cell adhesion and metastasis. Secreted proteins and peptides can be used as circulating tumor markers for diagnosis and monitoring

Polysomes translating membrane bound or secreted proteins are bound to the rough endoplasmic reticulum and can be separated from free cytosolic polysomes producing cytosolic proteins by sucrose gradient centrifugation. RNA from these two pool were hybridized to Affymetrix Genechips and the relative enrichment of each probeset within the MS or Cytoplasmic pool is reflected by the MS/CYT expression ratio. A training set of proteins with known location was obtained from Swissprot. 10-fold cross validation was used on the set of

genes with known annotated localization in order to determine ideal thresholds for total expression and MS/CYT ratio to maximize specifity.

755 probe sets were predicted membrane-associated or secreted, of which 432 had no previous subcellular location annotation. Based on the results of the 10-fold cross validation, it is likely a great number of the predicted MS genes will have MS localization. This is reflected by the average 97% positive predictive value observed in the 10-fold cross validation. Second, we examined the tentative annotations of genes in the set that were not used in the cross validation test and for which we predicted subcellular localization. Many of these have some tentative annotation which we do not consider definitive. Nevertheless, our MS predictions coincide with these tentative annotations 70% of the time.

Our Bayesian analysis may be over or under estimating MS localization, however, due to some violations of the equation assumptions. The localization of different genes are not entirely independent observations. For instance, there are clearly genes which co-localize due to genetic interactions. In addition, we make the assumption that these two classes are mutually exclusive which may not be true for a small fraction of genes. The RMA algorithm might be a different source of under-estimation for MS prediction, as it utilizes quantile normalization and might be over-correcting for underrepresented MS genes. It is possible that alternative microarray processing algorithms may yield additional predicted MS genes. Despite these drawbacks, we believe this will be a useful tool for investigators wishing to filter existing or future breast cancer Affymetrix datasets in order to look for MS genes. Alternative statistical methods may be useful for further analysis and confirmation of our results.

There are a significant number of genes with unambiguous MS annotation that fall below our MS/CYT threshold. It is unclear if this is due to a real biological process (some of those MS genes are not MS localized in MCF-7 cells, for instance) or a processing artifact. Further experimental analysis is needed to elucidate the mechanism in action. Further study is also needed to determine if the protein localization we discovered for MCF-7 cells holds true when analyzing other breast cancer cells.

References

1. van 't Veer, L.J., et al., *Gene expression profiling predicts clinical outcome of breast cancer.* Nature, 2002. **415**(6871): p. 530-6.